Speech Prosody 2016
31 May - 3 Jun 2106, Boston, USA

Non-referential gestures in adult and child speech: Are they prosodic?

Stefanie Shattuck-Hufnagel [1], *Ada Ren*[1], *Mili Mathew*[2], *Ivan Yuen*[3], *Katherine Demuth*[3]

[1] Speech Communication Group, RLE, MIT, USA
[2] Dr. S. R. Chandrasekhar Institute of Speech and Hearing, India, [3]Macquarie University, Australia
sshuf@mit.edu, peilin.ren@gmail.com, milimarym@gmail.com, ivan.yuen@mq.edu.au, katherine.demuth@mq.edu.au

Abstract

The manual gestures that accompany speaking have been analysed in terms of their form, their meaning, their role in the communicative act, and their timing with respect to the speech they accompany. Several schemes for categorizing these co-speech movements have been proposed, e.g. McNeill's (1992) iconic, metaphoric, deictic and beat gestures, and Kendon's (1994) distinction between substantive and pragmatic gestures. Among McNeill's gesture categories, *beats* are described as non-referential: simple flicks of the hand or finger, often performed repetitively and in rhythm, and lacking the complex phasing structure of other gesture types. For referential gestures, this complex phasing can include (in addition to the core stroke phase) preparation, pre- or post-stroke hold and recovery (Kendon 1980). Studies of adult speech show that many gestures are timed to overlap with phrase-level prosodic accents (Loehr 2004, Yasinnik et al. 2004); in at least one corpus of academic-lecture speech (Shattuck-Hufnagel et al. in prep.) these gestures are largely non-referential, like beats. We present evidence that this type of non-referential gesture can also have complex phasal structure in adults, and that children as young as 6 have such gestures in their repertoire, although less skillfully produced. Potential relations between prosody and gesture are discussed.

Index Terms: co-speech gesture, prosody, discourse structure, gesture-speech alignment, gesture types

1. Introduction

Many spoken utterances are accompanied by gestures that illustrate some aspect of the topic being spoken of; these gestures have been called *referential* (McNeill 1992), *imagistic* (i.e. illustrating as aspect of the message in visual terms, Kita 2001) and *substantive* (i.e. expressing utterance content, Kendon 1995). Such gestures, which include McNeill's (1992) iconic, metaphoric and deictic categories, illustrate in visual terms some aspect of the meaning of the message to be communicated. These meaning relationships include i) the description of an object expressed by a noun (e.g. a circling movement of an index finger extended downwards, in conjunction with the word 'cake'), and ii) the description of an action expressed by a verb (e.g. illustrating the act of mixing by a motion of the two hands), as well as iii) pointing to an object or location in space, and iv) a more abstract aspect, such as a framing gesture of the two hands when introducing a topic. In contrast to iconic, metaphoric and deictic gestures, which are referential, McNeill (1992) defines beats as *non-referential*, non-imagistic, i.e. without direct visual imagery relating to the spoken utterance, and suggests that they relate to the speech they accompany in other, possibly multiple, different ways. For example, he suggests that a beat may be a single simple flicking motion of the hand or finger, but also that beats, repeated in relatively rapid succession, often mark out the rhythm of the speech, keeping time much like a conductor marking the rhythm of a musical performance. This meaning is also reflected in Loehr's (2012) definition of beats as 'short rhythmic gestures used for emphasis' (p. 77). McNeill (1992) also raises the possibility that a beat may serve as a discourse marker, marking e.g. the introduction of a new character, summarizing the action, introducing new themes, etc. (p. 15). Thus, beats have been associated with focus marking, rhythmic marking, and discourse structure marking.

These multiple characteristics and functions assigned to gestural beats make it somewhat difficult to define this set of speech-accompanying gestures, a difficulty which may reflect a tendency to put all gestures which are not recognizably illustrative visually (i.e. are not metaphoric, iconic or deictic) into a single non-referential category. This issue is particularly clearly illustrated by an observation that emerged from an ongoing study of gestures produced by skilled academic lecturers (Shattuck-Hufnagel et al. 2007). Originally envisioned as a test of the hypothesis that gestures tend to co-occur with syllables that carry phrase-level prosodic prominences (pitch accents), this analysis revealed that by far the majority of the gestures produced by speakers in this context were non-referential. That is, they were not in any obvious way providing visual information about an object or an action, yet many of them seemed to have some of the additional gesture phases described as optional accompaniments to the strokes of referential gestures, i.e. preparation movements, pre- or pos-stroke holds, and recovery movements. This informal observation raised the question of whether non-referential gestures can be produced with complex phase structure, like that proposed by Kendon (1980), adopted by McNeill (2001, inter alia) and used by Loehr (2012) and others in their analyses of gesture form, structure and function. This paper describes some quantitative aspects of the phase structure of these apparently beat-like (i.e. non-representational) gestures, and reports the occurrence of similar gestures in a small pilot experiment involving 6-year-old children. Such findings raise interesting questions about the relationship between prosody and gesture, including the possibility that these two signalling systems share certain functions (e.g. marking the grouping of spoken elements, singling out a particular element for prominence, signalling the

 doi: 10.21437/SpeechProsody.2016-171

discourse function of an element), while differing in other ways. Thus in a sense they may serve as 'prosodic gestures'.

2. Evidence for adult use of non-referential gestures with phasal structure

As part of a larger study it was informally observed that many of the largely non-referential gestures produced by the speakers had additional phases accompanying their core stroke phase. To quantify this observation, a full phase analysis was carried out on the gestures included in this corpus.

2.1. Method

2.1.1. The Corpus. The sample of video recordings was drawn from commercially-available academic lectures. The advantages of these recordings include the visibility of the upper torso and arms of the speaker for most of the lecture (see Figure 1), and the extensive use of gesture by the speakers; one disadvantage is the interruption of the visual recording of the speaker by illustrative slides. While not a disadvantage for the present study, it may hamper any future study of the relationship of the gesture to discourse structure, since for some regions gestural analysis will not be possible.

Figure 1: Videos provide face-on views of the complete upper torso, including the hands, during most of the lecture.

The current sample includes videos of varying durations from 6 speakers, with 1988 stroke-defined gestures occurring in 46.5 minutes of speech. Non-gesticulatory movements were not included in the analysis (e.g. self-grooming movements, actions such as turning a page of notes or grasping the lectern, and small 'drifting' excursions that do not give the perceptual impression of intentional gesticulation).

Table 1

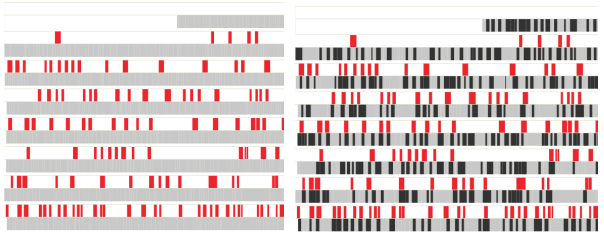
Speaker	Sample	#strokes	#min	strokes/min
Biology	1	58	0:01:08	51.18
Biology	2	42	0:01:00	42.00
Brettell	1and3	123	0:02:00	61.50
Brettell	2	51	0:01:10	43.71
Chaos	1	36	0:00:55	39.27
Chaos	2	25	0:01:12	20.83
Greece	1	53	0:01:14	42.97
Greece	2	48	0:01:15	38.40
HistEng	1	129	0:02:55	44.23
HistEng	2	89	0:03:10	28.11
London	all	1334	0:30:31	43.71
Total		**1988**	**0:46:30**	**42.753**

2.1.2. Labelling and Analysis

Stroke-defined gestures for all samples were labelled by one of the authors (AR) using ELAN to display the video (without listening to the sound) and record the labels. Optional phases accompanying each stroke, i.e. preparation, pre- or post-stroke hold and recovery, were labelled by trained labellers without listening to the sound. The separated sound file was used to label prosodic prominences and phrasing, carried out by the first author using Praat (Boersma and Weenink, 2014). Word/syllable alignments were estimated by a trained transcriber, also using Praat. Automatic computation of the overlap between strokes and accented syllables resulted in displays such as those in Figure 2.

Figure 2: Gesture strokes and pitch-accented syllables in the first 152 seconds of London lecture. Each horizontal line represents 19 seconds. Panel (a) shows gesture strokes as red vertical lines; panel (b) shows alignment with pitch-accented syllables (vertical black lines). Typically, gestural strokes occur often, and most (but not all) strokes co-occur with a pitch-accented syllable.

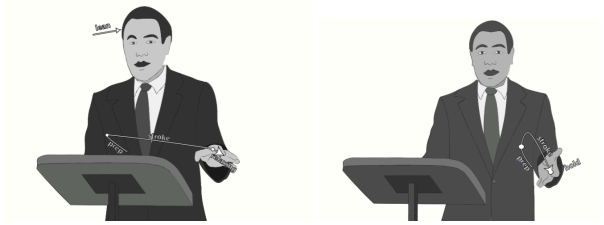
(a) (b)



In earlier work, Shattuck-Hufnagel et al. 2007) reported that in these lectures (as in other studies) gesture stroke end points are significantly likely to align with pitch accented syllables.

2.2. Results. A typical gesture in this corpus was non-referential, yet was realized with one or more of the optional phases (preparation, hold or recovery), illustrated in Figure 3.

Figure 3: Typical non-referential gestures with phasing structure.

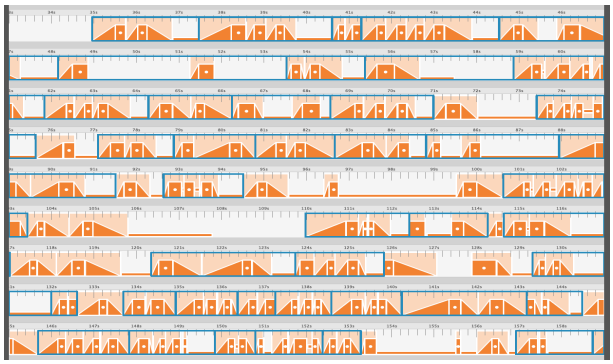


Quantitative results for the complex phasing of the stroke-defined gestures for the six speakers in this study are shown in Table 2, and a graphic illustration of the phasing is shown in Figure 4 (for the 2.5 minute sample in Figure 2). While the proportion of strokes with additional phases varies from speaker to speaker, it is clear that these non-representational gestures are often accompanied by one or more of the optional phases that have been described for representational gestures.

Table 2. The occurrence of optional phases with stroke-defined gestures (SDGs) for the 6 speakers analysed here.

Spkr	prep	pre-stroke hold	post-stroke hold	relaxing	relaxed	Total SDG
Biol	67	6	32	39	18	100
Bret	130	24	55	27	15	174
Chaos	49	0	19	21	12	61
Gree	86	0	9	45	33	101
HistEn	186	4	100	51	18	218
Lond	1134	5	46	483	327	1334

Figure 4: Time alignments of labelled phases of gestures in the sample shown in Fig 2. Polygons in this representation represent movement: Orange rectangles with white dots indicate strokes; a right triangle preceding a rectangle indicates a preparation; a right triangle following a rectangle indicates a recover. In contrast, horizontal lines represent non-movement: an orange line on the horizontal axis indicates a period when the hands are at rest. Rectangular blue outlines indicate perceived gesture groupings, which will not be discussed here. Note that most strokes have a preparation phase or recovery phase or both.

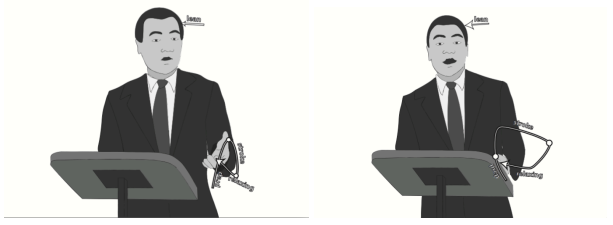


2.3. Discussion

The type of gesture observed in these academic lectures differs from the category of beats as described in the literature, in being non-referential but also having a stroke phase that is often accompanied by one or more of the optional phases of preparation, hold or recovery. While these movements are not overtly descriptive of the shape of an object referred to by a noun or the action referred to by a verb, as shown by the additional examples in Figure 5, they may well convey a more abstract meaning related to the structure of the discourse. This possibility has been suggested by a number of investigators, based on a small number of closely observed tokens. Investigation of the degree to which these non-representational gestures reflect such higher-level phenomena will require additional labelling of the discourse structure of these lectures.

Figure 5: (a) and (b) show two non-representational gestures, produced in quick succession, with the defining stroke phase accompanied by one or more additional phases; (c) illustrates a complex bimanual gesture by tracings of the location of the tip of the index finger of each hand, and (d) shows a simple uni-manual up-down movement (using the same method).

(a) (b)



(c) (d)

3. Non-referential gestures in children

Children as young as 2 years exhibit perceptual and production sensitivity to the prosodic structure of the ambient language they are learning (Demuth 1996). However, their use of various types of prosodic structure continues to develop until they reach adolescence (Wells et al. 2004). In the gestural domain, we have been interested to determine if and how 'prosodic' use of gesture may develop. However, as suggested above, it is not always clear from the previous literature what terminology has been used to refer to this type of co-speech gesture.

McNeill (1992) has suggested that 'beats', including finger flicks, are both abstract and discourse-referring. In comparing the multi-modal narrative abilities of five- and ten-year-old Italian, American and French children, Colletta et al. (2014) showed that beats occur less frequently in the narratives of young children than those of older children. However, it is not clear exactly what is meant by beats, and if any might be structured in the same way as seen in adults (cf. Shattuck-Hufnagel et. al., 2012).

We therefore wanted to investigate whether 6-year-old children might use gestures that are characterised as *an intentional movement that does not directly reflect contextual meaning, with a well-defined stroke phase.*

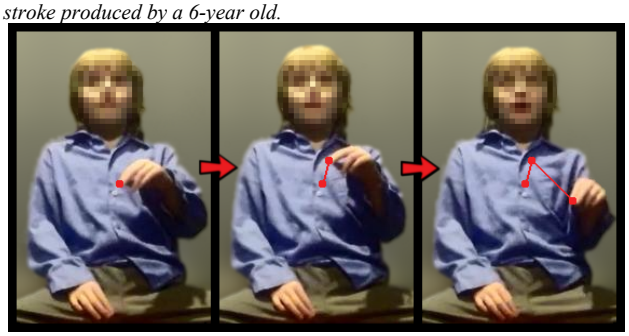
3.1 Method. Considering the previous literature, we hypothesized that children as young as 6 might employ this 'prosodic' use of gesture. Participants were Australian English-speaking children with no prior history of communication disorders. Nine children (5 boys, 4 girls) aged between 5;3 – 7;5 years (Mean 6;3 years) were included in the analysis. The children completed two tasks: a story narration after viewing a two minute movie clip, and an explanation task where the child planned a 'fantasy' family trip. Both tasks were carried out while interacting with their mothers, with the mother mostly playing the role of a facilitator. Both were also designed to avoid any specific need for deictic and/or referential use of gesture. The samples were coded for both *speech* (e.g., words, and number of turns) and *gestures* (*non prosodic* (iconic, metaphoric, deictic) vs. *prosodic*) in Praat (Boersma and Weenink, 2014) and ELAN respectively, by the third and fourth authors.

3.2 Results. The results showed that these 6-year-olds produced both referential and non-referential gestures, with the latter constituting 26% of all the gestures used. Forty-two percent of the non-referential gestures overlapped with a pitch accent, suggesting that this type of gesture is often used for emphasis and/or contrastive focus. However, a few were also observed during pauses, suggesting that this type of gesture is not always accompanied by speech. Phases of a non-referential gesture from one of the children are exemplified in Fig. 6.

A chi-square analysis was performed to determine whether use of gesture types varied as a function of discourse type (narrative vs. explanation). It did: χ^2 (1, N = 258) = 6.653, p = .01), with more use of referential gestures in the explanation task and more use of non-referential gestures in the narration task.

3.3 Discussion. These results suggest that 6-year-old English-speaking children, at least in Austrialia, use non-referential gestures with a well-defined stroke phase and additional phases, and that the use of these non-referential gestures can vary as a function of the discourse task. Furthermore, occurrence of these gestures was not restricted to the presence of pitch accent or even of speech, suggesting the possibility that these gestures might perform additional discourse functions not found in the adult lecturers, an interesting area for further research. Although these 6-year-olds were involved in a task very different from that of the adult lectures, and were not experts in the same way, they were still using gesture in a 'prosodic' fashion to communicate their intent to their respective audience.

Figure 6: Phases of a sample non-referential gesture: preparation & stroke produced by a 6-year old.



4. General Discussion and Conclusions

The definition of the type of speech-accompanying gesture termed a beat in the literature includes at least two types of events: single up-down or in-out 'flicks' of the hand or other articulator that indicate emphasis (or mark a particular aspect of discourse structure), and repeated productions of the same gesture in time with the rhythm of a spoken utterance. It is suggested that the complex phasing observed for representational gestures does not occur for beats. The present studies suggest the existence of a class of gestures which are not obviously representational, and therefore are candidates for designation as beats, yet often exhibit the kinds of complex phasing seen in representational gestures, particularly preparation phases. This result provides in initial step toward determining the forms and functions of non-representational gestures that co-occur with speech. The fact that so many of the gestures in these speech samples are non-representational is a bit surprising; it appears that different genres of speech can elicit different distributions of gesture types. Much gesture research has focussed on speech tasks with a substantial spatial component (e.g. the description of an apartment layout, or of the action of a cartoon), which may be particularly conducive to visually-illustrative gestures (Beatty & Shovelton 1999).

While the function of the type of gesture observed in these two studies is not yet clear, several authors have described a discourse-structure-signalling role for many types of speech-accompanying gestures, such as e.g. McNeill's (1992) cohesives, Kendon's (1980) locution-spanning gesture units, and Loehr's (2012) focus markers. In this sense, the functions of speech-accompanying gesture may overlap with those of spoken prosody, echoing the striking proposal in McNeill et al. (2001) that 'prosody is effectively gesture in spoken form'. The non-referential gestures observed in this study may be particularly likely to function like prosody in this sense, making them good candidates for 'prosodic gestures'.

In conjunction with earlier studies, these results raise interesting issues about the role of prosody and of gesture in speech production planning. McNeill and colleagues (2001 inter alia) and Kendon (1980 inter alia) have suggested that speech and gesture emerge together from a semantic/pragmatic plan for an utterance. For example, McNeill et al. (2001) note that 'motion, prosody and discourse structure are integrated at each moment of speaking' (p.9), and Kendon (1980) suggests that gesture and speech emerge as part of the same utterance plan. In a different cognitive framework, Keating & Shattuck-Hufnagel (2002) propose a prosodic planning structure for spoken utterances, while Turk & Shattuck-Hufnagel (2013) propose a phonetic planning module in which prosodic structure influences the surface phonetic word forms. The degree to which a prosodic planning frame governs the timing and form of both gesture and speech is a topic for future research.

Acknowledgements

We thank the gesture labellers supported by MIT's Undergraduate Research Opportunities Program, including Allison Mann and Cady Lytle, and ARC grant FL130100014 (to Demuth).

References

[1] Beattie, G. and Shovelton, H. (1999). *Journal of Language and Social Psychology* 18, 438-462

[2] Boersma, P., & Weenink, D. (2014). Praat: Doing Phonetics by Computer. Version 5.3.84. http://www.praat.org/

[3] Bolinger, D. (1986). *Intonation and its parts: Melody in spoken English*. Stanford, CA: Stanford University Press

[4] Colletta, J. M., Guidetti, M., Caprici, O., Cristilli, C., Demir, O. E., Kunene-Nicolas, R. N., & Levine, S. (2014). Effects of age and language on co-speech gesture production: an investigation of French, American, and Italian children's narratives. *Journal of Child Language*, FirstView Article, 1-24

[5] Demuth, K. (1996). The prosodic structure of early words. In J. Morgan & K. Demuth (eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*. Mahwah, N. J.: Lawrence Erlbaum Associates. pp. 171-184.

[6] Keating, P. & Shattuck-Hufnagel, S. (2002). A Prosodic View of Word Form Encoding. UCLA WPP, 112-156

[7] Kendon, A. (1980). Gesticulation and Speech: Two Aspects of the Process of Utterance. In M.R. Kay (ed.), *Nonverbal Communication and Language.* The Hague: Mouton. 207-227

[8] Kendon, A. (1995). Gestures as illocutionary and discourse structure markers. *J. Pragmatics* 23, 247-279

[9] Kita, S. (2001). Gesture in Linguistics. *International Encyclopedia of the Social & Behavioral Sciences,* 6215-6218

[10] Loehr, D. (2012). Temporal, structural and pragmatic synchrony between intonation and prosody. *J. Laboratory Phonol.* 3, 71-89

[11] McNeill, D. (1992). *Hand and Mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

[12] McNeill, D. (2001). Catchments, prosody & discourse. *Gesture* 1

[13] Shattuck-Hufnagel, S. et al. (2007). In Esposito et al. NATO vol.

[14] Shattuck-Hufnagel S., Ren P.L. (2012). Preliminaries to a Kinematics of Gestural Accents. Paper presented at the biannual conference of International Society for Gesture Studies, Lund

[15] Wells, B., Peppé, S., & Goulandris, N. (2004). Intonation development from 5 to 13. *J. Child Language*, 31, 749-778.

[16] Turk, A.E. and Shattuck-Hufnagel, S. (2013). What is speech rhythm? *Journal of Laboratory Phonology* 4, 93-118

[17] Yassinik, Y., Renwick, M., Shattuck-Hufnagel, S., 2004. The timing of speech-accompanying gestures. *JASA* 115, 2397